

## **RAW SEQUENCE LISTING**

**The Biotechnology Systems Branch of the Scientific and Technical  
Information Center (STIC) no errors detected.**

Application Serial Number: 10/549,944  
Source: TFWP  
Date Processed by STIC: 12/01/2006

# ***ENTERED***



IFWP

## RAW SEQUENCE LISTING

DATE: 12/01/2006

PATENT APPLICATION: US/10/549,944

TIME: 14:17:00

Input Set : A:\798\_2\_PCT\_SeqListing.TXT

Output Set: N:\CRF4\12012006\J549944.raw

```

4 <110> APPLICANT: Genencor International, Inc.
5     Jones, Brian E.
6     Grant, William D.
7     Heaphy, Shaun
8     Grant, Susan
10 <120> TITLE OF INVENTION: Novel Bacillus BagCel Cellulase
13 <130> FILE REFERENCE: GC798-2-PCT
C--> 15 <140> CURRENT APPLICATION NUMBER: US/10/549,944
C--> 16 <141> CURRENT FILING DATE: 2005-09-20
18 <150> PRIOR APPLICATION NUMBER: US 60/467,255
19 <151> PRIOR FILING DATE: 2003-04-30
21 <160> NUMBER OF SEQ ID NOS: 3
23 <170> SOFTWARE: FastSEQ for Windows Version 4.0
25 <210> SEQ ID NO: 1
26 <211> LENGTH: 4205
27 <212> TYPE: DNA
28 <213> ORGANISM: Bacillus sp.
30 <220> FEATURE:
31 <221> NAME/KEY: misc_feature
32 <222> LOCATION: (1)...(4205)
33 <223> OTHER INFORMATION: isolated from environmental water sample from
34     Sonachi Lake, Kenya
37 <400> SEQUENCE: 1
38 tctcatttag tttgactgaa atacctactg tgccacaaag taaagttaaa ctgacgaatg      60
39 tggaacgaat cacttaatgg ttctagcata gataacgaaa gatgaacacg ttcaaagttc      120
40 gccactcttt tgaaagaggg tgaacttttt ttgtgacaag aaaggggtgtt aaatgaagat      180
41 caaagctaaa caagatgagt aacggtttctt ttctttttta tagagtgagt tagtatatga      240
42 gtcccttata aatttctaga ctgttatttt aaataattga atgactcagt caccattaag      300
43 ttttcaacac ccataagcga cgtttgaaga tctaattgat cgagagggttt tatcactttg      360
44 gagcggaaga tcactgtagg actcgtttta tatggtgaac ttggtgttat tgtgtattta      420
45 aaagaaaggg aaacgaaaaa aatggttaaa ttagaaagag gctattacag agaggagaac      480
46 aaacaatgaa cgtaacactt gaagtgcacat actgcacgac taaaggtatt cgaacaacct      540
47 ttcattcaga aggtatggag gccgaaaaag caattacatc cgcagaagat tttcagcgga      600
48 caggacggat aaaacagatc gtctttagag atgagcgtga tagtccgtgg acgttaaaag      660
49 aacttaaaag atttttagaa gagattaaaa cggagccgca tcactctctc gtgtattttg      720
50 atgggggatt tgatttggag acacaacgat ctgggtcttg gtgtgattta ttatgaacaa      780
51 aatgacacgt cttatcgggt gagaagaaac gctaccgtgg cgtcattgac atcgaataac      840
52 gaagcagaat atgccgcttt acatttagga cttaaagaac ttgaagggat cgggtgcgcat      900
53 catctaccta tcactattta cgggtgattct caagttgtga tcaatcagtt aaaaggagaa      960
54 tgggctgtga tggaggaggt gttaaataaa tgggctgacc gtattgatca gcatttagct      1020
55 aaattaggca tgaccgctac ttataagtta atcccccgta aagaaaaccg tgaagcagat      1080
56 caactggcta cacaagcggt aaacgggcaa gaaattataa gtcaacgtga tgtcagtgag      1140
57 cgtggtgcag attagtctgc acccgcataa aagttaacgt atatagaagt ggatggggat      1200

```

## RAW SEQUENCE LISTING

DATE: 12/01/2006

PATENT APPLICATION: US/10/549,944

TIME: 14:17:00

Input Set : A:\798\_2\_PCT\_SeqListing.TXT

Output Set: N:\CRF4\12012006\J549944.raw

58	taaaggaacg	tcattcactc	taagcaagcg	ttgcgacagc	aaaaaagaaa	catataaggt	1260
59	ttttctgagc	tactatctat	acaaatagcc	aagtggcagt	taagctctta	cctcatcaag	1320
60	tttttgacta	ccagtcctcc	actcctactt	tcacattat	aaattgggtc	cttttttggt	1380
61	aataatcact	aattttgacg	gtatttttta	atagaaatat	atgctagatt	ataaaactagt	1440
62	aacgatgtag	aaggtggtga	ttgaccatat	aagaagactc	tttcaaacct	ggtagtatcg	1500
63	cattaaaaaa	tttgaaaggt	ggagaggaca	catgggttat	accaaagcga	agtgtacgtt	1560
64	gaaaaaaact	gtcttggttg	gtttaattct	ctgtttaagt	gtgtcaatgt	ttgttccaat	1620
65	gacatcagct	gaagatgtca	cttcgtcaca	gttggtatatt	cactcctatg	tagctgacat	1680
66	gcagcctggc	tggaaatttag	gaaatacgtt	tgacgctgtt	ggagatgatg	aaacagcgtg	1740
67	ggggaatcct	cgtgtaacaa	gagagttaat	aaaaacgatt	gctgatgaag	ggtataaaaag	1800
68	cattcgtatc	ccagtgcacat	ggcaaaatca	aatgggtggg	tctccagatt	atacgataaa	1860
69	tgaagattat	atcaatcggg	tggagcaagc	gatagattgg	gcgttgaggg	aagacttata	1920
70	tgtgatgtta	aatgtgcatc	atgactcatg	gctgtggatg	tatgatattg	aacataacta	1980
71	tgatgaggtc	atggcaagat	atacagctat	ttgggaacaa	ttgtcggaaa	aattcaaaaag	2040
72	ccactcccat	aagttgatgt	ttgagagtgt	caatgagcct	aggtttacgc	aggagtgggg	2100
73	agagattcaa	gaaaatcatc	atgcttactt	agaagattta	aataagacgt	tctattatat	2160
74	tgtcagagag	tcaggaggca	ataatgtgga	gcgcccttta	gtattgccta	cgatagaaac	2220
75	agccacgtct	caggatttac	tagatcgctt	gtatcaaaca	atggaagact	tggatgatcc	2280
76	ttattttaatt	gccacgggtgc	attattatgg	cttctggcca	tttagtgtca	atatagcagg	2340
77	tgactactcat	tttgaacagg	aaacacaaca	agatattata	gaacaccttg	accgtgttca	2400
78	taacacattt	acagcgcgtg	gtgtcccagt	tgtattaggg	gaattcgggt	tgtttaggctt	2460
79	tgacaaaagt	acggatgtga	ttcagcaagg	ggagaaatta	aagttttttg	agtttctcat	2520
80	ccatcatctc	aatgaacgtg	atataaccca	tatgttatgg	gataacggcc	agcatttaaa	2580
81	tcgagaaact	tatgcatggg	atgatcaaga	atttcatgac	atattaaaag	cgagttggga	2640
82	ggggcgcttc	gctacagcag	agtctaattt	gattcatgtg	aaggacggaa	agccaattag	2700
83	agatcaagat	atacagcttt	acttaaacgg	aaatgagcta	acagccttac	aggcagggga	2760
84	ggaatcgctt	gttctaggag	aggattatga	actagcagga	ggcgtattaa	cgctaaaagc	2820
85	ggacaccctc	acaagactaa	ttaccctcgg	tcaattagga	accaatgcag	tcatcacagc	2880
86	acaatttaatt	tctggagcag	actggcgctt	tcaattacag	aatgtggacg	tgccaacggg	2940
87	cgaaaataca	gatggctcaa	catggcattt	tgcgatccct	accattttta	atggtgatag	3000
88	tcttgcgacg	atggaagctg	tttatgcaaa	cggagaatat	gctggggccg	aagattggac	3060
89	gtcattttaa	gaatttggtg	aggcgctttc	tcctaattac	gccacagggg	aaattattat	3120
90	atcagaagcc	ttctttaacg	cggtagcgga	tgatgatata	catttaacat	ttcatttttg	3180
91	gagcggagag	acgggtggaat	ataccttacg	taaaaatggc	aattatgttc	aaggtagacg	3240
92	gtaacatgat	tttaattaat	agataaacca	gcctacctat	cgtttttgga	agaaggcaaa	3300
93	cgaatctcat	cttaccaca	ccgtgcttta	gaactttaga	agtgcgggtg	ttttttaaga	3360
94	catgaggaga	agcaatcctc	tatcaacagt	caccaatttt	tattcaggag	gtgtcaagtt	3420
95	atctaacgtt	ctatgaatgc	atatagtctc	tgacgaataa	acatagttaa	aaagaagtga	3480
96	gcctagtccc	cgaggggaag	gggataatgc	caacgtattg	gattaaagta	ccttcttgat	3540
97	aaaaagaaag	ggttttcaag	agatggaaat	gggctcgctt	gttatacttt	aattacgcct	3600
98	tggaacgtca	ttttggcggt	gctatttagt	aagatggctg	acatcataaa	agaggagtgg	3660
99	gttcgatggc	tttaattcaa	ttaagcttta	aatcacgagc	attaatgttg	caaacctctg	3720
100	tcaatgtttt	attaccgggtg	ggaatgaatg	ctgtagattt	tacaccaagt	gatgattttt	3780
101	cttatgttac	tgaccctttt	cctgtcctat	atcttttgca	tggtgcaact	gatgattatt	3840
102	cagcatggct	acgtctgtcc	tctattgaac	gatatgctga	agaaaaaaa	ttggcggtcg	3900
103	tcatgccaaa	tgctgatatg	agtgcgtata	cggataggtt	acatggccat	cgttactgga	3960
104	cgtatattag	taaggtgctg	cctgagttta	tgagagcaac	ttttcctatt	tctcagcacc	4020
105	gtgaagacac	ctttgcagct	ggtctgtcta	tgggaggata	cggggctttt	aaattggcgc	4080
106	tgcggaacc	ggaacgcttc	gctgcagctg	tgtcattatc	tggtgcagtt	gatatgagag	4140

## RAW SEQUENCE LISTING

DATE: 12/01/2006

PATENT APPLICATION: US/10/549,944

TIME: 14:17:00

Input Set : A:\798\_2\_PCT\_SeqListing.TXT

Output Set: N:\CRF4\12012006\J549944.raw

```

107 aagcaagtca accagactcc ctatttgtaa atgcctttgg tgaaggacg aaaatcgag 4200
108 ggaca 4205
110 <210> SEQ ID NO: 2
111 <211> LENGTH: 1713
112 <212> TYPE: DNA
113 <213> ORGANISM: Bacillus sp.
115 <220> FEATURE:
116 <221> NAME/KEY: misc_feature
117 <222> LOCATION: (1)...(1713)
118 <223> OTHER INFORMATION: isolated from environmental water sample from
119 Sonachi Lake, Kenya
122 <400> SEQUENCE: 2
123 atgggttata ccaaagcgaa gtgtacgttg aaaaaaactg tcttgtttgg ttttaattctc 60
124 tgtttaagtg tgtcaatgtt tgttccaatg acatcagctg aagatgtcac ttcgtcacag 120
125 ttggatattc actcctatgt agctgacatg cagcctggct ggaatttagg aaatacgttt 180
126 gacgctgttg gagatgatga aacagcgtgg gggaatcctc gtgtaacaag agagttaata 240
127 aaaacgattg ctgatgaagg gtataaaagc attcgtatcc cagtgcacatg gcaaaatcaa 300
128 atgggtgggt ctccagatta tacgataaat gaagattata tcaatcgggt ggagcaagcg 360
129 atagattggg cgttggagga agacttatat gtgatgttaa atgtgcatca tgactcatgg 420
130 ctgtggatgt atgatatgga acataactat gatgagggtca tggcaagata tacagctatt 480
131 tgggaacaat tgtcggaaaa attcaaaagc cactcccata agttgatgtt tgagagtgtc 540
132 aatgagccta ggtttacgca ggagtgggga gagattcaag aaaatcatca tgcttactta 600
133 gaagatttaa ataagacgtt ctattatatt gtcagagagt caggaggcaa taatgtggag 660
134 cgccctttag tattgcctac gatagaaaca gccacgtctc aggatttact agatcgcttg 720
135 tatcaaacia tggaagactt ggatgatcct tatttaattg ccacggtgca ttattatggc 780
136 ttctggccat ttagtgtcaa tatagcaggg tacactcatt ttgaacagga aacacaacia 840
137 gatattatag acacctttga ccgtgttcat aacacattta cagcgcgtgg tgtcccagtt 900
138 gtattaggcg aattcggttt gttaggcttt gacaaaagta cggatgtgat tcagcaaggg 960
139 gagaaattaa agttttttga gtttctcatc catcatctca atgaacgtga tataacccat 1020
140 atgttatggg ataacggcca gcatttaaatt cgagaaactt atgcatggta tgatcaagaa 1080
141 tttcatgaca tattaaaagc gagttgggag gggcgttctg ctacagcaga gtctaatttg 1140
142 attcatgtga aggacggaaa gccaattaga gatcaagata tacagcttta cttaaaggga 1200
143 aatgagctaa cagccttaca ggcaggggag gaatcgcttg ttctaggaga ggattatgaa 1260
144 ctagcaggag gcgtattaac gctaaaagcg gacacctca caagactaat taccctgggt 1320
145 caattaggaa ccaatgcagt catcacagca caatttaatt ctggagcaga ctggcgtttt 1380
146 caattacaga atgtggacgt gccaacggtc gaaaatacag atggctcaac atggcatttt 1440
147 gcgatcccta cccattttta tgggtgatgt cttgcgacga tggaaagctgt ttatgcaaac 1500
148 ggagaatatg ctgggcccga agattggacg tcatttaaag aatttggcga ggcgttttct 1560
149 cctaattacg ccacagggga aattattata tcagaagcct tctttaacgc ggtacgggat 1620
150 gatgatatcc atttaacatt tcatttttgg agcggagaga cgggtggaata taccttacgt 1680
151 aaaaatggca attatgttca aggtagacgg taa 1713
153 <210> SEQ ID NO: 3
154 <211> LENGTH: 570
155 <212> TYPE: PRT
156 <213> ORGANISM: Bacillus sp.
158 <220> FEATURE:
159 <221> NAME/KEY: VARIANT
160 <222> LOCATION: (1)...(570)
161 <223> OTHER INFORMATION: isolated from environmental water sample from

```

## RAW SEQUENCE LISTING

DATE: 12/01/2006

PATENT APPLICATION: US/10/549,944

TIME: 14:17:00

Input Set : A:\798\_2\_PCT\_SeqListing.TXT

Output Set: N:\CRF4\12012006\J549944.raw

162 Sonachi Lake, Kenya

165 &lt;400&gt; SEQUENCE: 3

```

166 Met Gly Tyr Thr Lys Ala Lys Cys Thr Leu Lys Lys Thr Val Leu Phe
167 1 5 10 15
168 Gly Leu Ile Leu Cys Leu Ser Val Ser Met Phe Val Pro Met Thr Ser
169 20 25 30
170 Ala Glu Asp Val Thr Ser Ser Gln Leu Asp Ile His Ser Tyr Val Ala
171 35 40 45
172 Asp Met Gln Pro Gly Trp Asn Leu Gly Asn Thr Phe Asp Ala Val Gly
173 50 55 60
174 Asp Asp Glu Thr Ala Trp Gly Asn Pro Arg Val Thr Arg Glu Leu Ile
175 65 70 75 80
176 Lys Thr Ile Ala Asp Glu Gly Tyr Lys Ser Ile Arg Ile Pro Val Thr
177 85 90 95
178 Trp Gln Asn Gln Met Gly Gly Ser Pro Asp Tyr Thr Ile Asn Glu Asp
179 100 105 110
180 Tyr Ile Asn Arg Val Glu Gln Ala Ile Asp Trp Ala Leu Glu Glu Asp
181 115 120 125
182 Leu Tyr Val Met Leu Asn Val His His Asp Ser Trp Leu Trp Met Tyr
183 130 135 140
184 Asp Met Glu His Asn Tyr Asp Glu Val Met Ala Arg Tyr Thr Ala Ile
185 145 150 155 160
186 Trp Glu Gln Leu Ser Glu Lys Phe Lys Ser His Ser His Lys Leu Met
187 165 170 175
188 Phe Glu Ser Val Asn Glu Pro Arg Phe Thr Gln Glu Trp Gly Glu Ile
189 180 185 190
190 Gln Glu Asn His His Ala Tyr Leu Glu Asp Leu Asn Lys Thr Phe Tyr
191 195 200 205
192 Tyr Ile Val Arg Glu Ser Gly Gly Asn Asn Val Glu Arg Pro Leu Val
193 210 215 220
194 Leu Pro Thr Ile Glu Thr Ala Thr Ser Gln Asp Leu Leu Asp Arg Leu
195 225 230 235 240
196 Tyr Gln Thr Met Glu Asp Leu Asp Asp Pro Tyr Leu Ile Ala Thr Val
197 245 250 255
198 His Tyr Tyr Gly Phe Trp Pro Phe Ser Val Asn Ile Ala Gly Tyr Thr
199 260 265 270
200 His Phe Glu Gln Glu Thr Gln Gln Asp Ile Ile Asp Thr Phe Asp Arg
201 275 280 285
202 Val His Asn Thr Phe Thr Ala Arg Gly Val Pro Val Val Leu Gly Glu
203 290 295 300
204 Phe Gly Leu Leu Gly Phe Asp Lys Ser Thr Asp Val Ile Gln Gln Gly
205 305 310 315 320
206 Glu Lys Leu Lys Phe Phe Glu Phe Leu Ile His His Leu Asn Glu Arg
207 325 330 335
208 Asp Ile Thr His Met Leu Trp Asp Asn Gly Gln His Leu Asn Arg Glu
209 340 345 350
210 Thr Tyr Ala Trp Tyr Asp Gln Glu Phe His Asp Ile Leu Lys Ala Ser
211 355 360 365
212 Trp Glu Gly Arg Ser Ala Thr Ala Glu Ser Asn Leu Ile His Val Lys

```

## RAW SEQUENCE LISTING

DATE: 12/01/2006

PATENT APPLICATION: US/10/549,944

TIME: 14:17:00

Input Set : A:\798\_2\_PCT\_SeqListing.TXT

Output Set: N:\CRF4\12012006\J549944.raw

```

213      370      375      380
214 Asp Gly Lys Pro Ile Arg Asp Gln Asp Ile Gln Leu Tyr Leu Asn Gly
215 385      390      395      400
216 Asn Glu Leu Thr Ala Leu Gln Ala Gly Glu Glu Ser Leu Val Leu Gly
217      405      410      415
218 Glu Asp Tyr Glu Leu Ala Gly Gly Val Leu Thr Leu Lys Ala Asp Thr
219      420      425      430
220 Leu Thr Arg Leu Ile Thr Pro Gly Gln Leu Gly Thr Asn Ala Val Ile
221      435      440      445
222 Thr Ala Gln Phe Asn Ser Gly Ala Asp Trp Arg Phe Gln Leu Gln Asn
223      450      455      460
224 Val Asp Val Pro Thr Val Glu Asn Thr Asp Gly Ser Thr Trp His Phe
225 465      470      475      480
226 Ala Ile Pro Thr His Phe Asn Gly Asp Ser Leu Ala Thr Met Glu Ala
227      485      490      495
228 Val Tyr Ala Asn Gly Glu Tyr Ala Gly Pro Gln Asp Trp Thr Ser Phe
229      500      505      510
230 Lys Glu Phe Gly Glu Ala Phe Ser Pro Asn Tyr Ala Thr Gly Glu Ile
231      515      520      525
232 Ile Ile Ser Glu Ala Phe Phe Asn Ala Val Arg Asp Asp Asp Ile His
233      530      535      540
234 Leu Thr Phe His Phe Trp Ser Gly Glu Thr Val Glu Tyr Thr Leu Arg
235 545      550      555      560
236 Lys Asn Gly Asn Tyr Val Gln Gly Arg Arg
237      565      570

```

**VERIFICATION SUMMARY**

PATENT APPLICATION: US/10/549,944

DATE: 12/01/2006

TIME: 14:17:01

Input Set : A:\798\_2\_PCT\_SeqListing.TXT

Output Set: N:\CRF4\12012006\J549944.raw

L:15 M:270 C: Current Application Number differs, Replaced Current Application Number

L:16 M:271 C: Current Filing Date differs, Replaced Current Filing Date